



О СПЕКТРАЛЬНОМ АНАЛИЗЕ ФОНЕМ С ИСПОЛЬЗОВАНИЕМ ЗВУКОВЫХ РЕДАКТОРОВ

И.А. СИДОРЕНКО
П.А. КУСКОВА

*Белгородский
государственный
национальный
исследовательский
университет*

e-mail:
Sidorenko@bsu.edu.ru

В статье рассмотрены причины вычисления искаженного спектра фонем русской речи при использовании звуковых редакторов. Сделан вывод о невозможности использования в учебном процессе звуковых редакторов для демонстрации формантной структуры фонем. Определены условия корректного вычисления спектра фонем.

Ключевые слова: фонема, форманта, звуковой редактор, быстрое преобразование Фурье.

Речевой обмен был и остается самым естественным и востребованным средством общения между людьми, поэтому разработка технологий обработки речевых сигналов всегда была и будет в центре внимания специалистов при создании перспективных инфокоммуникационных систем. Сферы применения речи обширны: это и обычный обмен информацией, и речевой ввод команд в компьютерных системах, синтез речи при автоматическом озвучивании текстовых файлов, системы автоматического создания текстовых документов на основе распознавания речи. В основе технологий применяемых для разработки подобных систем используются различные подходы, в том числе основанные на анализе фонемной и формантной структур речи [1, 3]. Известно, что фонемы, являясь, по сути, неделимыми элементами речи, характеризуются уникальным распределением энергии в спектральной области. В спектре у каждой фонемы присутствуют области повышенной концентрации энергии, называемые формантами. Расположение формант, их число и параметры позволяют однозначно установить фонему, поскольку определяются самим процессом порождения звуков голосовой системой человека. Изменяемый объем полостей гортани, рта, движения артикуляционного аппарата говорящего всегда подобны при создании каждого из звуков речи. Анатомические отличия людей друг от друга порождают отличия в высоте и тембре голоса, но характер распределение формант у фонем остается постоянным. Возможные отличия параметров голоса у людей вызывают смещения формантных областей по частоте, но не приводят к таким изменениям, чтобы одна фонема трансформировалась в другую. Фонемы, произнесенные разными людьми можно сравнить с различными почерками при написании текста на одном и том же языке – буква «а», независимо от того круглая она или с наклоном, всегда распознается как «а», а не как «е».

При подготовке специалистов в области инфокоммуникационных технологий важно обеспечивать наглядность процесса обучения, поэтому как при изучении природы самой речи, так и её элементов возникает потребность в демонстрации изучаемых понятий. Для этих целей логично применить звуковые редакторы, являющиеся многофункциональными прикладными программами, позволяющими в реальном масштабе времени производить различного рода манипуляции как с любыми звуковыми сигналами вообще, так и с речевыми сигналами в частности. Примером такого подхода к обучению могут служить практикумы с использованием программы Cool Edit Pro (в настоящее время купленной фирмой Adobe и известной под названием Adobe Audition), предлагаемые для повторения учащимися в учебном пособии [2]. Рекомендованные авторами пособия упражнения существенно повышают качество процесса обучения в сфере технологий цифровой обработки звука, однако при исследовании формантной структуры речи предлагают ограничиться только изучением влияния формант на разборчивость речи. Сами форманты, как объективные признаки фонем, для анализа и изучения не предлагаются. В связи с этим возникла потребность оценить возможность применения звуковых редакто-

ров для изучения формантной структуры звуков речи. В первую очередь, хотелось бы визуально убедиться в наличии и частотной локализации формантных областей.

Интенсивные исследования фонемной и формантной структуры речи проводились в середине прошлого века с использованием аналоговых приборов. В результате этих исследований были определены количество и типовые области частотной локализации формант для каждой фонемы, разработаны фонемные и формантные аналоговые вокодеры. Формантному синтезу речи посвящена, например, статья Савчука В.Л. [3], в которой приводится рисунок (см. ниже) и таблица с частотами формант для трех звуков русской речи.

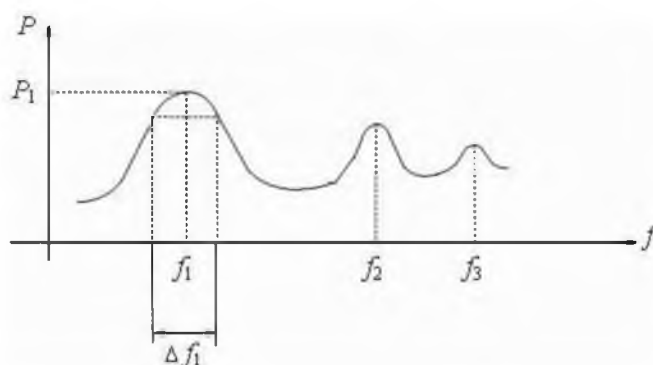


Рис. 1. Пример частотного спектра речи из статьи Савчука В.Л.[3]

Таблица 1

Таблица формантных частот из статьи Савчука В.Л. [3]

Фонема	Формантные частоты		
	F1	F2	F3
О	275	850	2400
И	250	2300	3000
А	575	900	2450

Таким образом, если осуществить микрофонную запись звуков речи (фонем), а затем с помощью звукового редактора выполнить их спектральный анализ, то должно быть возможным визуально убедиться в наличии формантных областей.

Для проведения исследований были выбраны три редактора звука. Это два наиболее известных и профессиональных звуковых редактора Adobe Audition 3.0 (вместе с его старой версией Cool Edit Pro 2.0) и Sound Forge 6.0. Третьим был выбран свободно распространяемый редактор Audacity. Авторам данной статьи были известны проблемы связанные с применением дискретного преобразования Фурье и алгоритма его практической реализации – быстрой преобразованием Фурье (БПФ). Об этом предупреждают и авторы пособия [2]. Тем не менее, была надежда на то, что удастся получить приемлемые для учебного процесса результаты. Однако на практике всё оказалось намного сложнее, чем ожидалось.

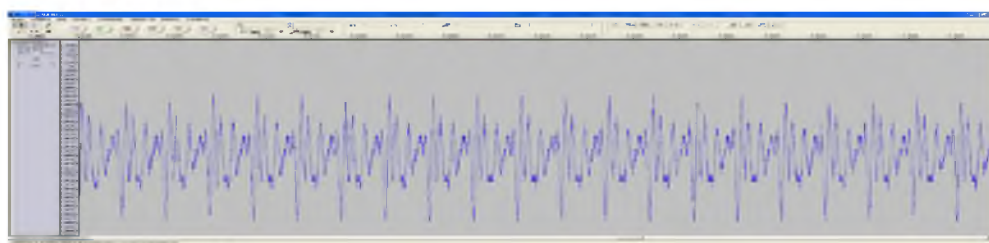


Рис. 2. Уровнеграмма долгого звука «а»

Из всего множества фонем была выбрана фонема, соответствующая звуку «а», поскольку гласные звуки имеют наиболее ярко выраженные формантные области в спектре фонем. В начале была выполнена микрофонная запись соответствующего звука с частотой дискретизации, принятой для речевых сигналов – 8 кГц и 8-ми битным квантованием. Уровнеграмма записанного звука приведена на рис. 2, на котором легко видеть периодически повторяющуюся фонему.



Однако попытка получить спектр данной фонемы не увенчалась успехом – получаемое изображение не соответствовало ожидаемому и имело одну формантную область в области низких частот до 1200 Гц. Для анализа выбиралось от одного до пяти и более периодов сигнала, но это не способствовало лучшему отображению спектра фонемы.

Анализ причин получения неудачного результата привел к следующим предположениям. Во-первых, при вычислении спектра фонемы не возможно точно согласовать объем выборки сигнала с параметрами окон БПФ преобразования, размеры которых должны быть кратны целой степени числа 2. Так, например, одна фонема записанного звука «а» при частоте дискретизации 8 кГц содержит порядка 38 отсчетов, две фонемы соответственно около 76, а минимальные размеры окна в звуковых редакторах – 64 или 128. Поэтому, в любом случае, анализируемый фрагмент при выполнении БПФ будет либо дополняться нулями, либо делиться на сегменты, равные размеру окна. Во-вторых, анализ фрагментов сигнала со скачками напряжения в точках разрыва и последующим умножением на функцию окна Хэннинга (или любую другую) для устранения эффекта Гиббса [4], само по себе всегда выступает источником искажений спектра.

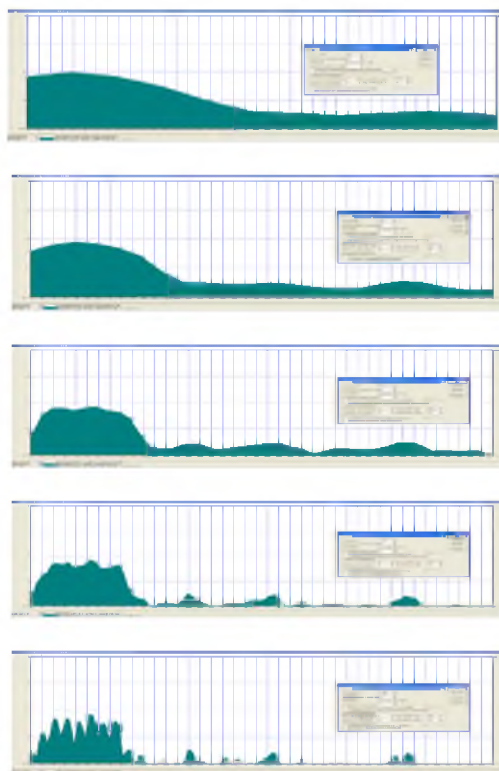


Рис. 3. Спектры фонемы звука «а» в программе Sound Forge 6.0 для размера окон БПФ 64, 128, 256, 1024, 2048

Далее, для полноты эксперимента была осуществлена запись звука «а» с частотой дискретизации 96 кГц и 16-ти битным квантованием и полученный сигнал анализировался с варьированием размеров окон БПФ, видом оконных функций, различным числом периодов фонем. Предполагалось, что увеличение числа отсчетов на длительности одной фонемы позволит повысить точность формантного анализа. Однако принципиально ничего не изменилось – получить изображение ожидаемых формантных областей не удалось. Изменение вида оконной функции и числа, выделяемых для анализа периодов записанного сигнала не существенно влияло на вид получаемого спектра. Увеличение размера окна БПФ, напротив, оказывает сильное влияние, но при этом попытка детализировать вид спектра приводит к отрицательному результату – спектр вырождается в гребенчатый с периодом частоты основного тона порядка 209 Гц, соответствующей, по видимому, частоте основного тона речи. Для примера на рисунке 3 показан вид спектра фонемы, полученный в программе Sound Forge 6.0 для размеров окон БПФ 64, 128, 256, 1024, 2048 при перекрытии 50% и взвешивающей функции окна Блэкмана-Харриса.

Имеющиеся на рисунке области локализации энергии в частотной области не соответствуют данным табл. 1, поэтому их детальное описание не приводится. При необходимости это можно увидеть, увеличив масштаб рисунка.

На рисунке 4 приведен спектр, вычисленный в программе Audacity с параметрами: размер окна БПФ 2048, весовая функция Хэмминга.

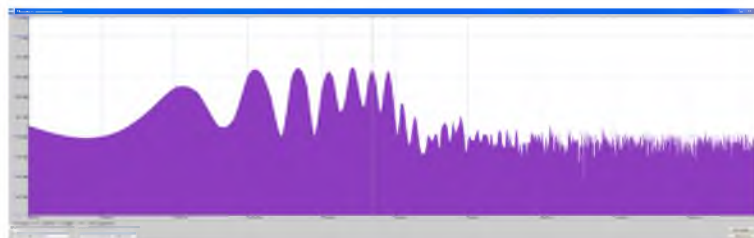


Рис. 4. Спектр фонемы звука «а» в программе Audacity для размера окна БПФ 2048

Аналогичные результаты получаются и в звуковом редакторе Adobe Audition 3.0. Для примера на рис. 5 показан фрагмент выделяемой фонемы длительностью 454 отсчета и вычисленный спектр при размере окна БПФ 128 отсчетов.



Рис. 5 Спектр фонемы звука «а» в программе Adobe Audition 3.0.
для размера окна БПФ 128 отсчетов

Из рисунка видно, что полученный спектр имеет флуктуационный вид, в котором трудно выделить формантные области за исключением единственной в области самых низких частот.

Таким образом, следует признать тот факт, что использование в учебном процессе звуковых редакторов для демонстрации формантной структуры фонем речи невозможно, поскольку вычисляемые ими спектры не отражают известных результатов. Для решения поставленной задачи следует использовать программное обеспечение, предоставляющее пользователю полную свободу в выборе параметров размера выборки анализируемого фрагмента и вычисляющего спектр по алгоритмам, не предусматривающим обязательную кратность размера окна БПФ степени числа 2. Такие условия могут быть реализованы, например, в программе MATLAB или в любой другой среде, дающей возможность произвольного задания параметров дискретного преобразования Фурье.



Список литературы

1. Медведев М.С. Использование вейвлет-преобразования для построения моделей фонем русского языка. // Вестник КрасГУ. Сер. Физ.-мат. Науки. 2006. Вып. 9. С.193-201.
2. Попов О.Б., Рихтер С.Г. Цифровая обработка сигналов в трактах звукового вещания: учеб. пособие для вузов. – М.: Горячая линия – Телеком, 2007.- 341 с.:ил.
3. Савчук В.Л. Формантный синтез. <http://www.ie.tusur.ru/books/COI/index.htm>.
4. Афонский А.А., Дьяконов В.П. Цифровые анализаторы спектра, сигналов и логики. Сер. «Библиотека инженера». – М.: СОЛОН-ПРЕСС, 2009. – 248 с.: ил.
5. Частотный анализ речевых сигналов. Жилияков Е.Г., Прохоренко Е.И. Научные ведомости Белгородского государственного университета. Серия: информатика и прикладная математика. – Белгород, 2006. – №2(31), выпуск 3. С. 201-208.

ON THE SPECTRAL ANALYSIS OF PHONEMES USING SOUND EDITORS

I.A. SIDORENKO

P.A. KUSKOVA

*Belgorod National
Research University*

*e-mail:
Sidorenko@bsu.edu.ru*

The article discusses the reasons for calculating the distorted spectrum of Russian speech phonemes using audio editors. It is concluded that the inability to use in the learning process sound editors to demonstrate the formant structure of phonemes. The conditions for the correct calculation of the spectrum of phonemes.

Keywords: phoneme formant, sound editor, fast Fourier transform.